Week 7: Activity 2

Tuesday's Recap

- Linear Algebra
- Linear Model Overview
- Simulating Data in R
- Fitting Linear Models in R
- Bayesian Inference

Today's Key Concepts

- Multivariate Normal distribution
- Conditional Multivariate normal distribution
- Gaussian Process Intro

Multivariate Normal Distribution

Formally, our matrix notation has used a multivariate normal distribution.

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}, \tag{1}$$

where $\underline{\epsilon} \sim N(\underline{0}, \Sigma)$, which also implies $\underline{y} \sim N(X\underline{\beta}, \Sigma)$.

Simulate data from a multivariate normal distribution. Use the x sequence created below and define $\Sigma_{ij} = \exp(-d_{ij})$ where Σ_{ij} is the i^{th} row and j^{th} column of Σ and d_{ij} is the distance between the i^{th} and j^{th} points.

```
x <- seq(0, 10, by = .1)
```

Partitioned Matrices

Now consider splitting the sampling units into two partitions such that $\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix}$. Then,

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \end{bmatrix} \sim N\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \underline{\beta}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} \right)$$

Fundamentally, there is no change to the model, we have just created "groups" by partitioning the model. Do note that Σ_{11} is an $n_1 \times n_1$ covariance matrix.

$$\Sigma_{11} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n_1} \\ \sigma_{22} & \sigma_2^2 & \cdots & \sigma_{2n_1} \\ \sigma_{31} & \sigma_{32} & \ddots & \sigma_{3n_1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_1 1} & \sigma_{n_1 2} & \ddots & \sigma_{n_1}^2 \end{bmatrix}$$

However, while $\Sigma_{12} = \Sigma_{21}^T$, neither of these are necessarily symmetric matrices. They also do not have any variance components, but rather just covariance terms. Σ_{12} will be an $n_1 \times n_2$ matrix.

$$\Sigma_{11} = \begin{bmatrix} \sigma_{1,n_1+1} & \sigma_{1,n_1+2} & \cdots & \sigma_{1,n_1+n_2} \\ \sigma_{2,n_1+1} & \sigma_{2,n1+2} & \cdots & \sigma_{2,n_1+n_2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_1,n_1+1} & \sigma_{n_1,n_1+2} & \ddots & \sigma_{n_1,n_1+n_2} \end{bmatrix}$$

Conditional Multivariate Normal

Here is where the magic happens with correlated data. Let $\underline{y_1}|\underline{Y_2} = \underline{y_2}$ be a conditional distribution for $\underline{y_1}$ given that $\underline{y_2}$ is known. Then

$$\underline{y_1}|\underline{y_2} \sim N\left(X_1\beta + \Sigma_{12}\Sigma_{22}^{-1}\left(\underline{y_2} - X_2\beta\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Now let's consider a few special cases (in the context of the DC housing dataset.) What is $\underline{y_1}|\underline{y_2} \sim$

1. Let $\Sigma = \sigma^2 I$, then the batch of houses in group 1 are conditionally dependent from the houses in group 2

2. Otherwise, let $\Sigma = \sigma^2 H$ and we'll assume Σ_{12} has some non-zero elements.

First a quick interlude about matrix inversion. The inverse of a symmetric matrix is defined such that $E \times E^{-1} = I$. We can calculate the inverse of a matrix for a 1×1 matrix, perhaps as 2×2, matrix and maybe even a 3×3 matrix. However, beyond that it is quite challenging and time consuming. Furthermore, it is also (relatively) time intensive for your computer.

Visual Example

Let $n_1 = 1$ and $n_2 = 1$, then

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right)$$

and

$$y_1|y_2 \sim N\left(\mu_1 + \sigma_{12}(\sigma_2^2)^{-1}\left(y_2 - \mu_2\right), \sigma_1^2 - \sigma_{12}(\sigma_2^2)^{-1}\sigma_{21}\right)$$

Now consider an illustration for a couple simple scenarios. Let $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$. Now assume $y_2 = -2$ and we compare the conditional distribution for a few values of $\sigma_{12} = \{0, .2, .8\}$. Plot these three distributions on the same figure.

Marginal Distributions

One last note, the marginal distributions for any partition $\underline{y_1}$ are quite simple.

$$\underline{y_1} \sim N\left(X_1\beta, \Sigma_{11}\right)$$

or just

$$y_1 \sim N\left(X_1\beta, \sigma_1^2\right)$$

if y_1 is scalar.

GP Overview

Now let's extend this idea to a Gaussian Process (GP). There are two fundamental ideas to a GP.

1. Any finite set of realizations (say $\underline{y_2}$) has a multivariate normal distribution.

2. Conditional on a set of realizations, all other locations (say $\underline{y_1}$) have a conditional normal distribution characterized by the mean, and most importantly the covariance function. Note the dimension of $\underline{y_1}$ can actually be infinite, such as defined on the real line.

The big question is how to we estimate Σ_{12}?

Fundamental idea of spatial statistics is that things close together tend to be similar.

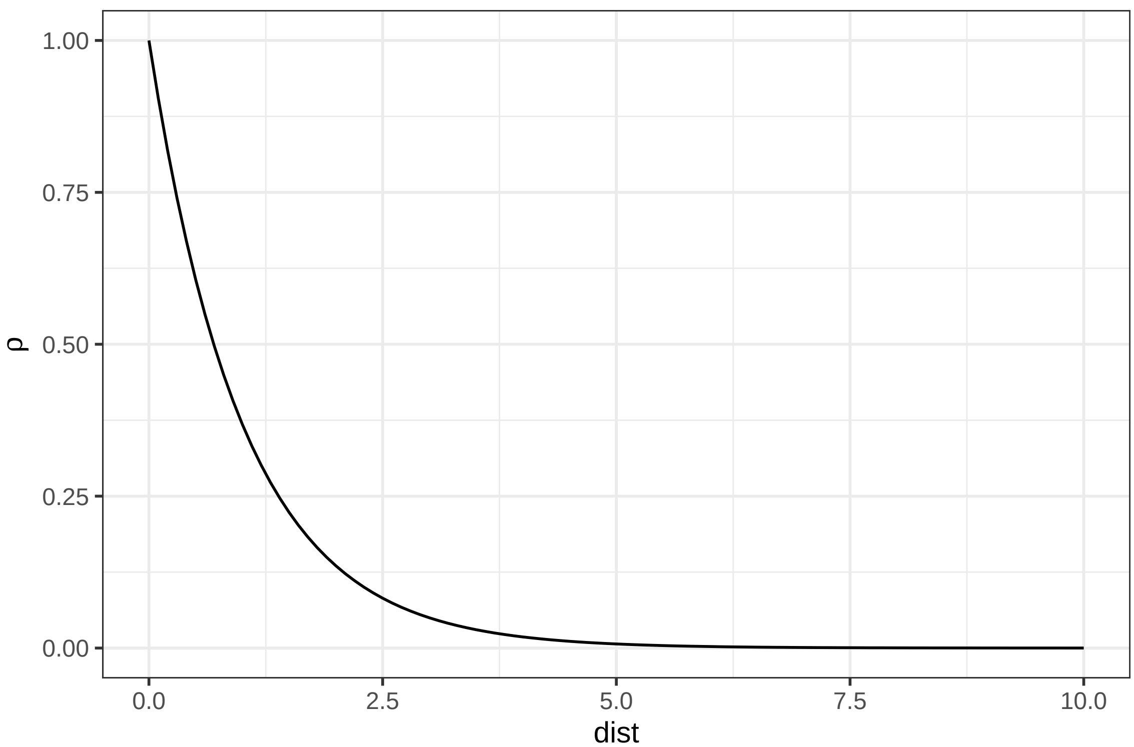
Correlation function

Initially, let's consider correlation as a function of distance, in one dimension or on a line.

As a starting point, consider a variant of what is known as the exponential covariance function - we used this earlier. First define d as the Euclidean distance between x_1 and x_2, such that $d = \sqrt{(x_i - x_j)^2}$

$$\rho_{i,j} = \exp(-d)$$

Lets view the exponential correlation as a function of distance between the two points.



Using a correlation function can reduce the number of unknown parameters in a covariance matrix. In an unrestricted case, Σ has $\binom{n}{2} + n$ unknown parameters. However, using a correlation function can reduce the number of unknown parameters substantially, generally less than 4.

Realizations of a Gaussian Process

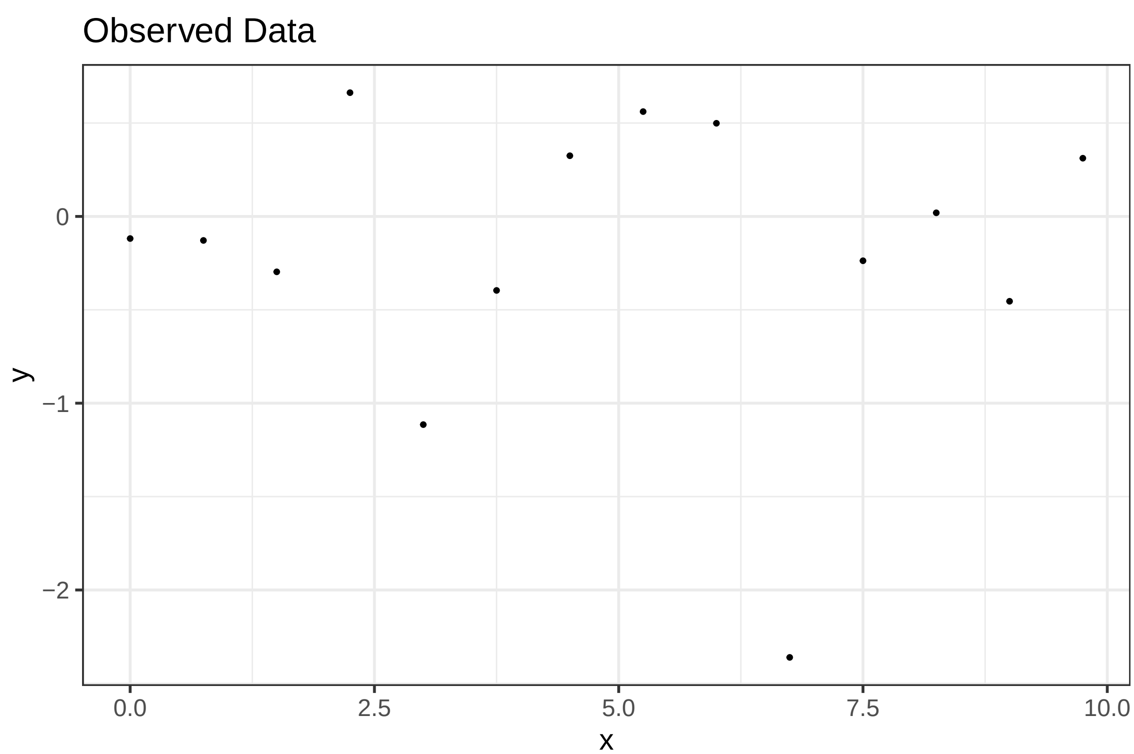
Recall that a process implies an infinite dimensional object. So we can generate a line rather than a discrete set of points. (While in practice the line will in fact be generated with a discrete set of points and then connected.)

For this scenario we will assume a zero-mean GP, with covariance equal to the correlation function using $\rho_{i,j} = \exp{(-d)}$

Simulate and visualize a realization from this process.

Connecting a GP to conditional normal

Now consider a discrete set of points, say $\underline{y_2}$, how can we estimate the response for the remainder of the values in the interval [0,1].



We can connect the dots (with uncertainty) using:

$$\underline{y_1}|\underline{y_2} \sim N\left(X_1\beta + \Sigma_{12}\Sigma_{22}^{-1}\left(\underline{y_2} - X_2\beta\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Plot the conditional mean curve on the observed data.

GP Regression

Now rather than specifying a zero-mean GP, let the mean be $X\underline{\beta}$.

```
x <- seq(0, 10, by = .25)
beta <- 1
n <- length(x)
d <- sqrt(plgp::distance(x))
H <- exp(-d)
y <- rmnorm(1, x * beta ,H)
```